



Evaluating Teaching with Multiple Measures

Laura Goe, Ph.D.

Teacher evaluation has many purposes, including accountability, tenure and promotion decisions, and identification of teachers in need of support. However, **the ultimate goal of teacher evaluation should be to improve teaching and learning.** By using multiple measures to evaluate instruction and student learning growth, and then aligning the results with professional growth opportunities and support, teaching and learning can be improved.

We know that teaching has a greater impact on student success than any other school-based factor. Research has shown huge differences among teachers in their abilities to help students achieve at high levels. These differences are not just *between* schools, but *within* schools, meaning that in most schools, you can find outstanding teachers whose students are learning more and learning at a faster rate. By the same token, you can find teachers in almost any school whose students are not making the expected progress. Providing support for those teachers to adjust instruction and improve student outcomes is clearly a priority, but helping *all* teachers reach their full potential is key to improving the profession.

Recognizing the importance of good teaching, federal priorities have focused on using student achievement as a “significant” measure of teacher effectiveness, and many states have recently passed legislation to include student achievement in their teacher evaluation systems. This shift in priorities has been motivated by the persistence of achievement gaps between poor and privileged students and between minority and white students. Too many students across the country are failing to reach proficiency, and they are dropping out of school at alarming rates. Clearly, more needs to be done to improve educational opportunities and outcomes, particularly for those students who are most at risk. For teachers, the rapid shift in priorities has raised many questions and concerns.

There is a growing emphasis on revamping current evaluation systems that are based almost exclusively on classroom observations. When classroom observa-

tions are connected to teaching standards, and when they can accurately place teachers on a continuum of performance, they can be valuable tools for identifying teachers whose students are likely to perform at high levels, as well as those teachers who may need support to become more effective. However, the relationship between teacher observations and student test scores depends not only on a good evaluation tool, but also on *trained* evaluators who are able to accurately distinguish among teacher performance levels.

Even if classroom observations are no longer the *only* measure of teacher performance, they are still a vitally important component of a comprehensive teacher evaluation system because they can help identify areas of a teacher’s practice that may need adjustment. *But unless we also know whether students are learning at appropriate levels, we will not know how much and what kinds of support and professional growth opportunities a teacher needs to make adjustments.* Some teachers may need new classroom management strategies to improve the learning environment for their students, while other teachers, who are already effective with most learners, may benefit from different instructional strategies so they can better serve the students who are struggling with content. And while some teachers may improve by simply making adjustments, others may require more intensive support. Using student learning growth to identify classrooms where student achievement is far below expectations is an important first step toward figuring out what strategies and supports are needed to improve learning.

Multiple measures and their uses

Student learning: Multiple measures of student learning capture important information about a teacher’s instruction, but also provide valuable information to teachers that can help them adjust their instructional focus and strategies. Standardized test scores provide strong evidence of student knowledge of specific content, but they occur late in the academic year and can only measure *growth* if they are compared

with those same students' results on previous tests of the same subject. Students' *growth* in knowledge and skills can be measured by starting with assessing what they know and can do at the beginning of the school year, then recording their progress through classroom-based tests (such as DIBELS); curriculum-based tests for subjects like psychology, history and business; products created by students (e.g., for art class or industrial arts); student performances (such as musical or dramatic performances); and portfolios that chronicle students' early and later efforts, showing improvement over time. By assessing student progress throughout the year, teachers not only document learning growth, they identify students who are in need of targeted assistance. If these students' needs are not identified until *after* standardized test results are in teachers' hands, the moment for learning important concepts in a particular subject may have passed.

There are many ways to assess what students know and can do in addition to standardized tests; using a variety of assessments can help teachers better understand students' progress in "real time," so they can target instruction appropriately. Furthermore, it is important for teachers, researchers and policymakers to better understand how various types of assessments:

- contribute to an understanding of student learning and how it can be measured fairly and efficiently;
- help clarify the relationships among various assessments; and
- show whether various assessments that are given in K-12 can predict later success in college and the workplace.

Instructional quality: Classroom observations are not very useful as a means of promoting teachers' knowledge and professional growth unless they are preceded by a conference discussing the context for the lesson, the learning objectives for students, and the instructional strategies the teacher will use. The observation should be followed by an opportunity for discussion and feedback. Peer assistance and review (PAR) is one program that capitalizes on expert teachers' knowledge of practice by training them to conduct classroom observations. PAR may be especially important to new or struggling teachers who need extra attention. PAR evaluators can provide detailed feedback, point teachers toward resources, and recommend instructional and classroom management strategies.

Besides observations, evaluation of teachers' lesson plans, student learning objectives, classroom/home-work assignments and student work may also show whether teachers have high expectations for student learning and are providing them with challenging assignments. Any of these evaluation methods should be accompanied by a conference in which the teacher can offer additional information and the evaluator can ask questions and provide feedback.

Professional responsibility and collaboration: Teachers contribute to school stability and success in many ways, as well as to the well-being and success of students—and not just their own students. Some observation instruments, such as Charlotte Danielson's Framework for Teaching and Praxis III, include professional responsibility as an important part of successful teaching. Teachers collect evidence to show their participation in school reform efforts, parent conferences, school events, and lesson study or other professional learning community activities, as well as their willingness to work with teachers of English language learners and special educators. Evaluations should also recognize teachers for collaborating and sharing responsibility for student learning through their participation in Response to Intervention teams or other teacher-led efforts to improve students' opportunities to succeed. Professional responsibility may be measured not only for individual teachers but also for teams or groups of teachers (such as teachers who share students in a certain grade, or teachers of a particular subject). Evaluating groups of teachers rather than individual teachers sends a clear message that teacher collaboration and mutual responsibility for student success are valued.

Why different types of evidence are needed for teacher evaluation

Multiple measures are good for both teachers and students: They provide more information about student learning that teachers can use to improve instruction, and more information about how teachers have contributed to student learning growth. Some of the specific benefits of using multiple measures to evaluate teaching are listed below. Multiple measures:

Strengthen teacher evaluation.

- They provide a more complete picture of a teacher's contribution to student learning.
- They contribute to greater confidence in the results of a teacher's evaluation.
- They provide more information about collaboration for students' success.

Contribute to teachers' professional growth.

- They create opportunities for teachers to learn from their colleagues.
- They provide teachers with greater insights into how their instruction is affecting student learning.

Set the stage for improved teaching and learning.

- They offer more complete evidence about students' learning growth, particularly in nontested subjects and grades.
- They provide more complete evidence of learning growth for English language learners and students with disabilities.
- They contribute to a more comprehensive view of students' strengths and of areas where they need improvement.

Strengthening the evaluation

More complete picture: Multiple measures provide a more complete picture of a teacher’s contribution to student learning. Here’s an example. Teacher A tailors her instruction narrowly to what she believes students need to know for the state standardized test, and spends long hours on test preparation. Teacher B focuses on meeting the content standards through rich, engaging lessons and activities that help students develop a deep understanding of the subject. Both teachers have good results on the state test. But which teacher’s classroom would you want *your* child in? Clearly, *how* teachers get good test results matters. We want to know that good student outcomes are the result of good teaching, not a narrow focus on what is tested.

Greater confidence in the results: Multiple measures also contribute to greater confidence in the results of a teacher’s evaluation. In the previous example, if multiple measures were used for her evaluation, Teacher A would likely have received a lower rating on instructional practice, since her efforts focused narrowly on tested content and on raising test scores. However, also based on multiple measures, Teacher B would have received a higher rating on instructional practice, because she engaged and challenged the students with lessons that made the content meaningful to them. Additional evidence from examining a sample of the teachers’ assignments would likely result in a higher overall rating for Teacher B, who gives interesting assignments that encourage deep thinking, rather than multiple-choice worksheets.

More information about collaboration: Multiple measures can provide more information about collaboration for students’ success. Teachers, counselors and specialized personnel frequently collaborate to help struggling students, meeting in “triage” or Response to Intervention teams. Working together to identify students’ needs and set learning objectives for individual students, they also coordinate efforts with the classroom teacher to monitor progress and adjust strategies. Such coordinated efforts keep students from being held back or referred to special education. If credit is given just to one teacher for learning growth and improved test scores, the important contributions of others who had a hand in those improvements go unrecognized. A comprehensive evaluation system should include ways to record the efforts of the staff who contribute to student success. Minutes or notes from team meetings, records of student assessment results, logs of interventions and the results of those interventions, and other relevant information can be collected and used for the evaluations of all the participating staff.

Contributing to professional growth

More opportunities to learn from colleagues: Multiple measures provide more opportunities for teachers to learn from their colleagues. In every school,

there are teachers whose students are achieving at higher levels than expected. Observations of those teachers reveal students engaged in challenging activities within a structured learning environment that encourages respect and cooperation. Those teachers’ lesson plans and learning objectives reveal high expectations for students and high standards for success. Such high-performing classrooms are ready-made learning labs for “discouraged” teachers, since they provide strong evidence that all students, even those with disadvantages, can learn at high levels given the right structure and opportunities. Before discouraged teachers can create that structure and those opportunities, they must first believe such success is possible. Then they need guidance and support to create similar learning environments in their own classrooms.

Greater insights on effectiveness of instructional practices: Multiple measures provide administrators and teachers with more information to improve teaching and learning. Scores on student achievement tests might be used by a teacher to determine that his students are doing fine with subject-verb agreement but struggling with punctuation. He may then set appropriate learning objectives for student mastery of punctuation. However, the test results do not provide the teacher with information about *the specific practices and instructional strategies* that might be used to improve students’ understanding of punctuation. Other measures that involve feedback from evaluators or colleagues—such as classroom observations and analysis of lesson plans, for example—might be more helpful in connecting teaching practices to the learning objectives he has set. Such additional measures also give the evaluator a chance to steer the teacher toward resources available in the school or district to support his professional growth.

Setting the stage for improved teaching and learning

More complete evidence about student learning growth, particularly in nontested subjects and grades: Multiple measures can provide more complete evidence about student learning growth, particularly in nontested subjects and grades. In a comprehensive teacher evaluation system, teachers of nontested subjects can demonstrate that their students are learning through student performances and projects, portfolios, curriculum-based tests, and classroom assessments performed by teachers. Projects or performances in subjects such as art, music, drama, physical education and industrial arts, for example, can be evaluated to show growth over time. Student progress toward meeting subject standards is usually documented with the grades students receive at the end of the course. However, to measure *teachers’ contributions to student growth*, it is necessary to determine what the student knows and can do at the beginning of the year. Districts and states around the country are currently exploring or piloting ways of measuring growth in these nontested subjects and

grades so that *all* teachers can be recognized for their contributions to student learning.

More evidence of learning growth for English language learners and students with disabilities: Multiple measures can provide more evidence of learning growth for English language learners and students with disabilities. Even though most students with special needs and English language learners participate in standardized testing, there are other important learning objectives for these students that cannot be measured with subject-specific tests. For example, it is important to measure progress on behavioral goals or nonacademic performance objectives for students with special needs. Measuring English language learners' progress toward mastery of English is also important. States and districts are considering ways to credit teachers both for the quality of Individual Education Plans (IEPs) for students with special needs and for their students' progress in meeting goals. They are also exploring ways to credit teachers with English language learners' progress toward English mastery.

Multiple measures in practice

Throughout the country, states and districts are working toward identifying ways to incorporate multiple measures into teacher evaluation systems. Delaware and Tennessee, successful Race to the Top applicants, have put together collaborative teams that include teachers, administrators and union representatives. These teams are working together to determine which measures to use and how to implement them successfully. The AFT's Innovation Grant efforts in New York and Rhode Island are demonstrating that teachers and administrators, doing their own research and collaborating with evaluation experts, can design comprehensive teacher development and evaluation systems that will be rigorous, fair, and focused on improving teaching and learning. These models will provide guidance and inspiration to other states and districts as they launch their own efforts in the coming years.

Dr. Goe is a former middle school teacher in special education and English/language arts who worked in high-poverty communities in Mississippi and Tennessee. She is a research scientist at Educational Testing Service and principal investigator for research and dissemination for the National Comprehensive Center for Teacher Quality. She currently advises a number of states and districts on creating comprehensive teacher evaluation systems that meet federal and state priorities.



A Union of Professionals

American Federation of Teachers, AFL-CIO
555 New Jersey Ave. N.W. • Washington, DC 20001 • 202/879-4400 • www.aft.org